MICROBIAL GENOMICS

SHORT COMMUNICATION

Parras-Moltó and Aguirre de Cárcer, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000466

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

A comprehensive human minimal gut metagenome extends the host's metabolic potential

Marcos Parras-Moltó and Daniel Aguirre de Cárcer*

Abstract

Accumulating evidence suggests that humans could be considered as holobionts in which the gut microbiota play essential functions. Initial metagenomic studies reported a pattern of shared genes in the gut microbiome of different individuals, leading to the definition of the minimal gut metagenome as the set of microbial genes necessary for homeostasis and present in all healthy individuals. This study analyses the minimal gut metagenome of the most comprehensive dataset available, including individuals from agriculturalist and industrialist societies, also embodying highly diverse ethnic and geographical backgrounds. The outcome, based on metagenomic predictions for community composition data, resulted in a minimal metagenome comprising 3412 genes, mapping to 1856 reactions and 128 metabolic pathways predicted to occur across all individuals. These results were substantiated by the analysis of two additional datasets describing the microbial community compositions of larger Western cohorts, as well as a substantial shotgun metagenomics dataset. Subsequent analyses showed the plausible metabolic complementarity provided by the minimal gut metagenome to the human genome.

DATA SUMMARY

The datasets analysed during the current study are available from their original sources; Global and Flemish datasets can be downloaded from MG-RAST and EBI under accession numbers QIIME:850 and EGAS00001001689, respectively. Core Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology identifiers, reactions and compounds are available within File S1 (available with the online version of this article). The scripts employed in this study have been deposited at GitHub (github.com/marcosparrasmolto/Picrust_analysis).

INTRODUCTION

The study of the human gut microbiome has drawn from different disciplines (e.g. microbiology, ecology, genomics), and has substantiated the idea that humans could be considered as holobionts [1] in which the gut microbiota play essential functions [2, 3]. Knowledge of what constitutes a healthy gut microbiome is regarded as pivotal [4] for the development of predictive models for the diagnosis and management of gut microbiome-related maladies. However, the strong inter-subject variability in community composition observed in cross-sectional studies [5] hindered an early definition of a set of bacterial species common to all healthy humans [6]. While recent efforts have been able to detect such a health-related set in terms of shared taxonomic assignments [4, 7], and more precisely in terms of shared 16S rRNA sequence clusters of varying phylogenetic depth [8] whose existence has been related to a shared intra-cluster ecology [9], the idea that a healthy gut microbiome 'core' may exist only in terms of function [10] remains widespread.

In this regard, early high-throughput shotgun metagenomic studies already reported a strong pattern of shared genes in the gut microbiome of different individuals [11, 12]. These results led to the definition of a novel concept, the minimal gut metagenome [12], defined as the set of microbial genes necessary for the homeostasis of the whole gut ecosystem, and expected to be present in all healthy humans. The idea that the gut microbiome provides a specific set of functionalities shared by all individuals is intuitive. However, it is still unclear whether these functionalities could arise from a shared set of genes or from different combinations of genes. Moreover, if the host were to play a greatly diminished role as a selective force on its resident gut microbiome, when compared to

Received 12 May 2020; Accepted 16 October 2020; Published 03 November 2020

Author affiliations: [1]Departamento de Biología, Universidad Autónoma de Madrid, Madrid 28049, Spain.

***Correspondence:** Daniel Aguirre de Cárcer, daniel.aguirre@uam.es

Keywords: 16S rRNA gene; community assembly; hologenome; human gut microbiome; metagenomics; PICRUSTs.

Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; KO, KEGG orthology; NSTI, nearest sequenced taxon index; OTU, operational taxonomic unit.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary figures and one supplementary file are available with the online version of this article.

external factors such as diet, then there would be no set of microbial functionalities shared by all humans. Nevertheless, despite its potential as a conceptual framework with which to study the gut ecosystem, the minimal gut metagenome concept has received little attention in the literature following its initial definition and description in terms of various ubiquitous metabolic pathways [10–12] and recent description of prevalent pathways in a larger cohort [13].

Hence, the aim of the present study was to recapitulate the minimal human gut metagenome conceptual framework and provide a proof-of-concept of its utility. More specifically, we set out to identify the 'core genes' (defined as the set of genes detected in all individuals), jointly comprising the minimal gut metagenome, as well as the 'core reactions' (defined as the set of metabolic reactions detected in all individuals). According to the minimal gut metagenome concept, the former should be related to gut homeostasis at large (i.e. not only metabolic homeostasis). However, knowledge on the latter should improve our understanding of the gut microbiome's ability to augment human metabolism.

For knowledge of the minimal gut metagenome to be most useful, it should pertain more to *Homo sapiens* as a species and, hence, should not be solely focused on Western cohorts. Unfortunately, most human gut shotgun metagenomic datasets are very restricted in terms of lifestyles and ethnicities, mostly arising from Western and (or) industrialist cohorts [10–14]. This study analysed the minimal gut metagenome on the basis of 16S rRNA gene-based metagenomic predictions from the most comprehensive dataset available (dataset *Global* – 382 individuals from rural Malawi, metropolitan USA and Venezuelan Amerindians [15]; see Table 1), which despite its comparatively smaller cohort size is far more inclusive in terms of global distribution, lifestyle and ethnicity, specifically including agriculturalist and industrialist societies from three continents. We then compared the *Global* dataset with two larger 16S rRNA datasets from Western cohorts (dataset *Flemish* – 873 individuals from Belgium [4]; and dataset *Twins* – 2727 individuals from the UK [16]), as well as to a substantial shotgun metagenomics dataset (dataset *Shotgun* – gene abundances from 123 individuals from the USA, Europe and China obtained from the 2017 paper by Bradley and Pollard [17]), and compared with the human genome to assess the degree to which the minimal metagenome may complement and expand its host's metabolic potential.

Impact Statement

The gut microbiome is essential to our wellbeing and, thus, knowledge of what constitutes a healthy microbiome is regarded as pivotal for the management of microbiome-related maladies. In this regard, the minimal gut metagenome is defined as the set of microbial genes necessary for the homeostasis of the whole ecosystem, and expected to be present in all healthy individuals. Despite its interest, it had received little attention following its initial definition and overview in Western cohorts. We analysed the minimal gut metagenome in an inclusive dataset in terms of global distribution, lifestyle and ethnicity, significantly including agriculturalist and industrialist societies from three continents. The results, browsable and substantiated with complementary datasets, now pertain more to our species, rather than to industrialist societies of particular ethnic and geographical backgrounds, and indicate that the human gut minimal metagenome could extensively contribute to the human holobiont's metabolic potential.

METHODS

Datasets

All datasets comprised 16S rRNA gene sequences obtained using primer pair F515–R806 targeting the V4 hypervariable region, with the exception of dataset *Shotgun*, which included Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (KO) identifier [18] abundances obtained through shotgun sequencing of metagenomic DNA [17]. All sequence data was derived from stool samples from healthy subjects over 3 years old, with no history of recent antibiotic treatment prior to sampling (see Table 1).

Metagenomic predictions

QIIME [19] scripts were employed during initial sequence processing (File S1). Briefly, datasets were independently

Table 1. Characteristics of the datasets

Name	Geographical distribution	No. of individuals	Sequence depth*	Read length (bp)	Sequencing technology (Illumina)
Global	Malawi, USA, Venezuela	382	>300,000	100	GAIIx
Twins	UK	2727	>15,000	2×250	MiSeq
Flemish	Belgium	873	>8,000	2×250	MiSeq
Shotgun	USA, Europe, China	123	15,000,000	2×75, 2×100	GAIIx, HiSeq

*Values represent final sequence depth per sample before analysis (i.e. after chimaera removal and subsampling to common depth).

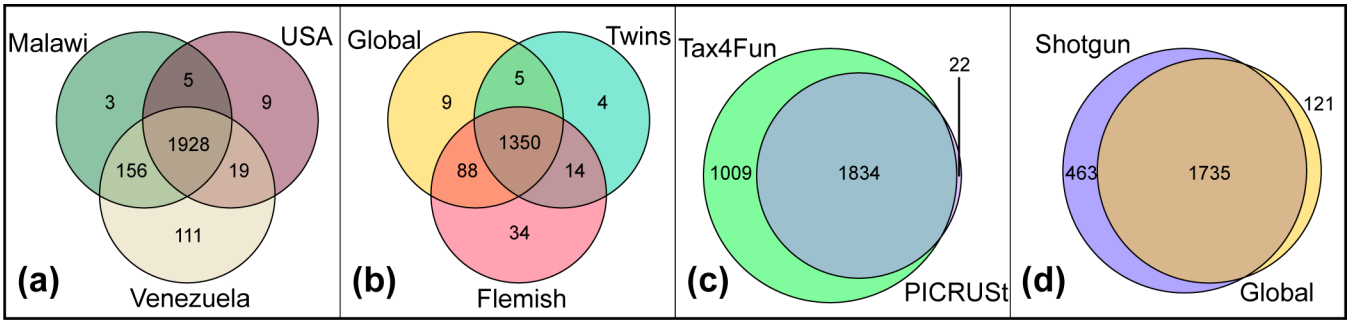

Fig. 1. Venn diagrams depicting the overlap in core reactions between different datasets and software. (a) Different sample sets within *Global*. Values refer to the analysis with the same number of individuals per population (50). (b) Different 16S rRNA datasets. Values refer to the analysis with the same number of sequences per sample (8000). (c) Differences between metagenomic prediction software. (d) Differences between 16S rRNA (*Global*) and shotgun metagenomics (*Shotgun*) datasets.

processed as follows. Firstly, they were subsampled to the minimum common depth. Then, chimeric sequences were identified with *usearch61* [20] and removed. Finally, sequences were clustered into operational taxonomic units (OTUs) using Greengenes [21] 0.97 representative sequence dataset (May 2013) as a reference using *usearch61*. Subsequently, PICRUSt scripts were employed to first normalize OTU abundances by 16S rRNA gene copy number, and then transform normalized OTU abundances into KO abundances. Correlation between predictions and measurements was evaluated using the nearest sequenced taxon index (NSTI; describes the novelty of organisms within a community with respect to previously sequenced genomes [22]) as a proxy for the Spearman coefficient, as they are strongly negatively and significantly correlated [22]. Tax4Fun [23], an alternative metagenome prediction pipeline, was also employed with the *Global* dataset following the suggested standard procedure.

Since more than one KO group may carry out a particular reaction, KO abundances were mapped to KEGG reactions. In cases where a KO identifier mapped to more than one reaction, all reactions linked to the KO were scored. KOs and reactions appearing in all individuals in the datasets were defined as core. Finally, the MinPath algorithm [24] was used for biological pathway reconstruction from core KOs.

Metabolic complementarity assessments

Host–microbiome cooperation was assessed with NetCooperate [25] using the metabolic complementarity index. This index provides a quantification of the extent to which two species may support one another through biosynthetic complementarity. There is no threshold for 'complementarity' and 'no complementarity' and, hence, the metrics have to be employed in a comparative manner [25]. Here, the index was used to study both moieties of the human holobiont: the human genome and the minimal gut metagenome. Hence, the reciprocal analysis evaluates the relative strength of each moiety complementing the other. To do so, core reactions were transformed into linked KEGG compounds, and then analysed with NetCooperate. To further assess such complementarity, both the core reactions and the reactions encoded by the human genome were imported into the interactive metabolic pathway explorer iPATH3.0 [26].

RESULTS

The results show that 5865 KO groups were predicted as present in *Global*'s pan-metagenome, while the minimal gut metagenome represented 3412 KOs (i.e. the core genes), which can in turn be mapped to 1856 reactions (i.e. core reactions) and 128 complete metabolic pathways (File S1). As could be expected, lowering the prevalence threshold used to define core reactions (100%) increased the number of core reactions, but mainly in a gentle-slope linear fashion (Fig. S1). The core metagenome was very similar among the three distinct sample sets comprising *Global* (Fig. 1a), with the USA set showing the smallest set of core reactions, and less overlap with Malawian and Venezuelan samples. However, *Global*'s core reaction set was comparatively similar to those obtained using Western-like datasets *Twins* and *Flemish* (Fig. 1b).

The presented core reactions were predicted from 16S rRNA profiles using an ancestral-state reconstruction algorithm (PICRUSt). However, the set of core reactions was substantiated by the use of Tax4fun [23], a taxonomy assignments-based approach (Fig. 1c). PICRUSt's predictions seem conservative (more appropriate for a minimum estimate, as intended) since they are a subset of Tax4fun predictions. More importantly, *Global*'s core reaction set presented a high overlap to that obtained from a substantial shotgun metagenomics dataset targeting the human gut microbiome [17], chosen among those publicly available based on the number of individuals and geographical and ethnic distribution (Fig. 1d). The 463 reactions described as core in *Shotgun* but not in *Global* (Fig. 1d) likely arise from the smaller size of the *Shotgun*'s cohort, as well as its increased lifestyle, environmental and genetic homogeneity (Table 1). However, the great majority of core reactions in *Global* not described as core in *Shotgun* still presented a very high prevalence in the dataset (Fig. S2); 1735 out of 1856 (93.5%) core reactions in *Global* are also core

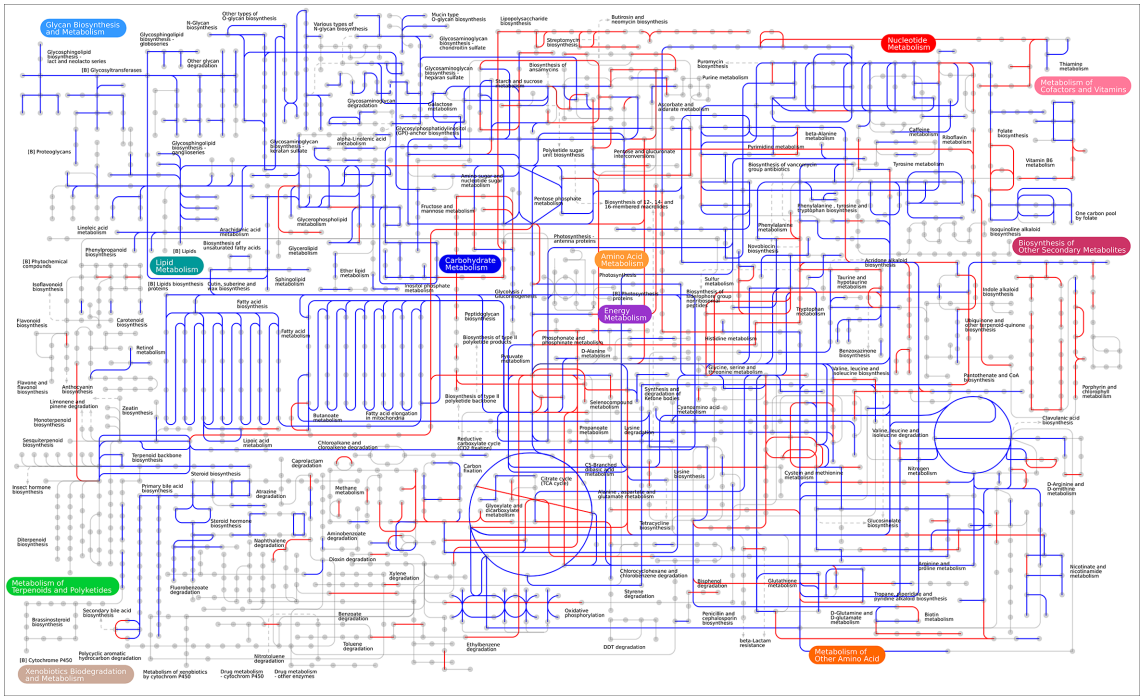

Fig. 2. The minimal gut metagenome extends human metabolic potential. Nodes in the map correspond to chemical compounds and edges represent enzymatic reactions. The figure provides an iPath3.0 representation of KEGG metabolic pathways, where reactions catalysed by enzymes encoded in the human genome appear in blue, while core reactions of the human gut pan-microbiome not encoded also by the human genome appear in red (see File S1 for the reaction list allowing web-based visualization and exploration).

reactions (100% prevalence) in *Shotgun*. Only 37 (2%) core reactions in *Global* had a prevalence level <95% in *Shotgun*, and 6 (0.32%) reactions had a prevalence level below 75%. No apparent shared functional or taxonomic origin affiliation was found for these six reactions. Within the *Global* dataset, there was a positive correlation between prevalence and mean abundance (Fig. S3). Nevertheless, while all core reactions featured relatively high mean abundance values, many similarly abundant reactions presented lower prevalence values.

In addition to providing an improved description of the human minimal gut metagenome, the present study aimed to assess its complementarity to the human genome. In this regard, the metabolic complementarity judged by the metabolic complementarity index [25] was >2 times larger when considering the human metabolism being complemented by *Global*'s minimal gut metagenome, when compared to the inverse (0.0807 and 0.0386, respectively).

Considering two metabolites as linked if they represent the substrate and product of a core reaction, within the overall metabolic map (Figs 2 and S4) 199 microbial metabolites link with 89 *H. sapien*s metabolites through 256 core reactions, representing the predicted extended metabolic capability of the human holobiont provided by its gut ecosystem. Additionally, the map pinpoints 55 core reactions and 84 metabolites with no apparent connection to *H. sapiens* metabolism, as well as 36 core reactions able to link *H. sapiens* metabolites by reactions different to those carried out by enzymes encoded within the human genome.

Not surprisingly, several core reactions are implicated in the production of short-chain fatty acids, such as butyrate and acetate, which are known to have an active role in normal human physiology (e.g. fuel for several cell types, regulation of gene expression, differentiation and inflammation) [27, 28]. Another hallmark of the predicted minimal gut metagenome relates to the presence of core reactions implicated in the production of several vitamins (B1, B2, B5, B6, B9, H, K1, K2, L1, coenzyme B12), several of which had previously been shown to be produced by common gut commensals [29].

DISCUSSION

Over the last decade, a large number of studies have generated shotgun metagenomic sequencing data from human faecal samples (for a significant list see the work by Nayfach *et al.* [30]). However, most samples analysed originated from diseased individuals, and to a lesser extent infants; thus, they are not fit for the purpose of studying the healthy human gut metagenome. Available suitable shotgun datasets comprising comparatively large cohorts present a narrow geographical distribution and originate from individuals living in industrialist societies, such as those from the Human Microbiome Project [11] (242 individuals, USA) and MetaHit consortia [14] (249 individuals, Europe). A noteworthy exception is that

reported by Brito *et al*. featuring 81 samples from individuals from an industrialist society (USA) and 172 individuals from a rural society (Fiji) [31]. Unfortunately, this second cohort was recruited regardless of health status and, thus, deemed as not suitable for the abovementioned purpose.

Indeed, the results presented herein are influenced by the fact that the metagenomic prediction approach employed is, to a certain extent, biased, as explained below (see Limitations section). As such, the core genes and reactions reported should be taken cautiously. Thus, validation of each particular core reaction in the ecosystem, as well as the possibility of each core metabolite traversing the membrane, along with its potential significance to the host, is beyond the scope of this study. Nevertheless, returning to the three possible scenarios of shared functionality in the human gut pan-microbiome postulated above, (i) no shared functionality, (ii) shared functionality related to different combinations of genes, and (iii) shared functionality related to a shared combination of genes, the results are strongly supportive of the latter. Thus, we believe that the minimal gut metagenome idea indeed represents a potentially useful conceptual framework with which to improve our knowledge of the role played by the human gut microbiome on maintaining host homeostasis.

According to our results, the human gut minimal metagenome could extensively contribute to the human metabolic potential. The browsable core reactions reported here (File S1) represent a highly restrictive set, since reactions need to be present in all subjects to achieve the core status. Using a different approach, Abu-Ali *et al*. recently described a core human gut metagenome in samples from 308 healthy individuals (65–81 year old men, USA) as 407 pathways with detectable DNA in at least two samples [32]. Most importantly, our core reactions were predicted as present in all subjects from a cohort including individuals from agriculturalist and industrialist societies, also embodying highly diverse genetic, ethnic and geographical backgrounds. Furthermore, the results were validated using additional large-cohort datasets, as well as a substantial shotgun metagenomics dataset. Hence, the described minimal gut metagenome now pertains more to *H. sapiens* as a species, rather than to industrialist societies of particular ethnic and geographical backgrounds. Finally, our results seem to indicate that the minimal metagenome has a greater role in complementing the human metabolism than the other way around.

The fact that the Venezuela samples presented the largest core reactions set could be attributed to its presumed more cohesive lifestyle. Interestingly, the results indicate that the USA population restricted the number of detected core reactions, since Venezuela and Malawian samples presented an additional 156 reactions with 100% prevalence in their joint dataset, compared to <20 exclusively shared with 100% prevalence between USA samples and any of the other groups. Moreover, these values may be conservative, since the reference genomes may be biased towards bacterial strains more frequent in industrialist countries. This reduction in functional overlap provides circumstantial support to the emerging concern that industrialist populations may have lost the microbial diversity needed to adequately sustain a healthy host [33].

The minimal gut metagenome was defined as the set of microbial genes necessary for the homeostasis of the whole gut ecosystem and present in all individuals [12]. Its essence overlaps with the hologenome concept: the combined genomes of host and associated microbiota [1]. Our results sustain the hologenome concept and, thus, the view of humans as holobionts, insofar as they are supportive of the existence of a set of microbial metabolic genes present in all individuals studied. However, our results cannot be used to discern whether the human host and its resident gut microbiota can be considered a single unit of evolution [34], the strictest holobiont/hologenome definition, or if these terms may only refer to a useful eco-evolutionary framework [35].

Limitations

In this study, 16S rRNA gene-based metagenomic predictions were employed in the assessment of the minimal human gut metagenome to be able to profit from the more comprehensive 16S rRNA datasets. These datasets greatly outclass available human gut shotgun metagenomic datasets in terms of cohort size, geographical distribution, ethnic and lifestyle diversity, and to a certain extent depth of sequencing. In a sense, one read in a shotgun metagenomics dataset represents one gene count, while one read in a 16S rRNA amplicon survey represents, via metagenomic prediction, one genome count. However, the use of metagenomic predictions presents various limitations and possible biases, which have been explored previously [22], the most noteworthy being that it only infers the bacterial and archaeal component of the metagenome, is significantly affected by both the quality of available genome annotations and the fact that available genomes are not evenly distributed across the phylogeny, the lack of perfect one-to-one mapping between genomes and even full-length 16S rRNA sequences, and primer bias. Nevertheless, the ability to count almost three orders of magnitude more genes in a metagenomic sample per sequence (with the number of bacterial genes per genome normally in the very few thousands), even as a prediction, is still useful.

Here, functional predictions based on 16S rRNA phylogenetic marker gene sequences were obtained using PICRUSt, a computational approach that has shown large and significant correlation in predicting metagenomic abundances from 16S rRNA measurements (Spearman r=0.82, P <0.001) and synthetic communities (Spearman r=0.9, P <0.001) [22]. To date, PICRUSt has been used in a myriad of scientific works and different research scenarios, such as the analysis of environmental samples [36], medically relevant communities [37] or *in vitro* assemblies [38]. The authors of the PICRUSt paper state that there is a significant negative correlation (Spearman r = −0.4, P <0.001)

between NSTI values and Spearman correlation between empirical shotgun metagenome abundances and PICRUSt predictions based on 16S rRNA sequences [22]. Here, NSTI values for the different sample sets of *Global* (0.135±0.021, 0.098±0.018 and 0.131±0.023 for Malawian, USA and Venezuelan samples, respectively; see Fig. S5) were lower (generally correlated with higher correlation between metagenomic measurements and 16S rRNA predictions) than those previously reported for soil samples (0.17±0.02), which showed a significant (P <0.001) correlation between predictions and matched shotgun metagenomics assignments [22]. Also, the more extreme NSTI values reported for the Human Microbiome Project dataset, with NSTI values ranging 0.10–0.15, still presented high correlation coefficients between metagenomic measurements and 16S rRNA predictions [22].

The NSTI values that we obtained for human gut microbiome samples fall within the range of NSTI values for samples in the PICRUSt validation that had high correlation between metagenomic abundance measurements and 16S rRNA predictions [22]. In this regard, an enhanced and updated report on the utility, correlation between predicted and experimental measurements, and accuracy of PICRUSt's predictions would be welcomed by the community, more so since this area of development seems to remain active [39, 40]. The values obtained were not homogenous among the three distinct sample sets in the *Global* dataset, with values for both the Venezuelan and Malawian samples being roughly 35% higher than that of the USA samples. In this regard, the detected functional overlap could be somewhat inflated, since the reference genome set employed is likely biased towards strains obtained from industrialist countries.

Funding information
This work was funded by the Spanish Ministry of Science and Innovation, grant numbers BIO2016-80101-R and PID2019-108797RB-I00.

Author contributions
D. A., conceptualization and writing. M. P. and D. A., formal analysis.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Bordenstein SR, Theis KR. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol* 2015;13:e1002226.
2. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science* 2005;307:1915–1920.
3. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ *et al*. Metagenomic analysis of the human distal gut microbiome. *Science* 2006;312:1355–1359.
4. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y *et al*. Population-level analysis of gut microbiome variation. *Science* 2016;352:560–564.
5. Aguirre de Cárcer D, Cuív PO, Wang T, Kang S, Worthley D *et al*. Numerical ecology validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria along the human colon. *ISME J* 2011;5:801–809.
6. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R *et al*. The human microbiome project. *Nature* 2007;449:804–810.
7. Zhang J, Guo Z, Xue Z, Sun Z, Zhang M *et al*. A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. *ISME J* 2015;9:1979–1990.
8. Aguirre de Cárcer D. The human gut pan-microbiome presents a compositional core formed by discrete phylogenetic units. *Sci Rep* 2018;8:14069.
9. Aguirre de Cárcer D. A conceptual framework for the phylogenetically constrained assembly of microbial communities. *Microbiome* 2019;7:142.
10. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A *et al*. A core gut microbiome in obese and lean twins. *Nature* 2009;457:480–484.
11. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–214.
12. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS *et al*. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464:59–65.
13. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J *et al*. Strains, functions and dynamics in the expanded human microbiome project. *Nature* 2017;550:61–66.
14. Li J, Jia H, Cai X, Zhong H, Feng Q *et al*. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 2014;32:834–841.
15. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG *et al*. Human gut microbiome viewed across age and geography. *Nature* 2012;486:222–227.
16. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R *et al*. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 2016;19:731–743.
17. Bradley PH, Pollard KS. Proteobacteria explain significant functional variability in the human gut microbiome. *Microbiome* 2017;5:36.
18. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016;44:D457–D462.
19. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–336.
20. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460–2461.
21. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL *et al*. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;72:5069–5072.
22. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D *et al*. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013;31:814–821.
23. Aßhauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 2015;31:2882–2884.
24. Ye Y, Doak TG. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* 2009;5:14.
25. Levy R, Carr R, Kreimer A, Freilich S, Borenstein E. NetCooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinformatics* 2015;16:164.
26. Darzi Y, Letunic I, Bork P, Yamada T. iPath3.0: interactive pathways explorer v3. *Nucleic Acids Res* 2018;46:W510–W513.
27. Donohoe DR, Garge N, Zhang X, Sun W, O'Connell TM *et al*. The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab* 2011;13:517–526.
28. Furusawa Y, Obata Y, Fukuda S, Endo TA, Nakato G *et al*. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 2013;504:446–450.

29. Letunic I, Yamada T, Kanehisa M, Bork P. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci* 2008;33:101–103.
30. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. *Nature* 2019;568:505–510.
31. Brito IL, Yilmaz S, Huang K, Xu L, Jupiter SD *et al*. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* 2016;535:435–439.
32. Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T *et al*. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat Microbiol* 2018;3:356–366.
33. Blaser MJ. The past and future biology of the human microbiome in an age of extinctions. *Cell* 2018;172:1173–1177.
34. Moran NA, Sloan DB. The hologenome concept: helpful or hollow? *PLoS Biol* 2015;13:e1002311.
35. Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF *et al*. Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiomes. *mSystems* 2016;1:e00028-16.
36. Bier RL, Voss KA, Bernhardt ES. Bacterial community responses to a gradient of alkaline mountaintop mine drainage in Central Appalachian streams. *ISME J* 2015;9:1378–1390.
37. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L *et al*. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* 2015;517:205–208.
38. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M *et al*. Emergent simplicity in microbial community assembly. *Science* 2018;361:469–474.
39. Douglas GM, Beiko RG, Langille MGI. Predicting the functional potential of the microbiome from marker genes using PICRUSt. *Methods Mol Biol* 2018;1849:169–177.
40. Douglas GM, Maffei VJ, Zaneveld J, Yurgel SN, Brown JR *et al*. PICRUSt2: an improved and extensible approach for metagenome inference. *bioRxiv* 2019;672295.